

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 02-224068

(43)Date of publication of application : 06.09.1990

(51)Int.Cl.

G06F 15/40

(21)Application number : 01-042940

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 27.02.1989

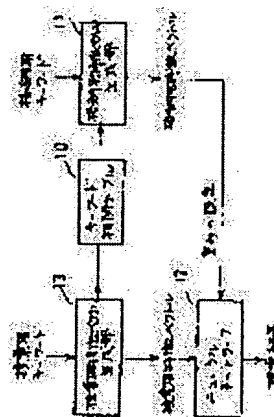
(72)Inventor : KOYANAGI SHIGERU

## (54) INFORMATION RETRIEVING SYSTEM

### (57)Abstract:

PURPOSE: To attain the retrieval based on a correlation between key words and to realize the retrieval of a high function by utilizing a neural network.

CONSTITUTION: As for data to be stored in a data base, a key work thereto is stored in a neural network 12 together with a feature vector transformed by a key word correlation table 10. That is, the feature vector of each stored data is held as a weight vector of a neurone corresponding to the stored data. Also, at the time of retrieval, with respect to a key work to be retrieved, a retrieval vector transformed by using the key word correlation table 10 is used, and the similarity of this vector and the feature vector of each stored data is derived by the neural network 12. In such a manner, the retrieval can utilize a function using the correlation between the key words more than mere key word matching, and a retrieval function which abounds in flexibility can be realized.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

⑫ 公開特許公報(A) 平2-224068

⑤ Int. Cl.<sup>3</sup>

G 06 F 15/40

識別記号

5 1 0 J

庁内整理番号

7313-5B

④ 公開 平成2年(1990)9月6日

審査請求 未請求 請求項の数 2 (全4頁)

⑥ 発明の名称 情報検索システム

⑦ 特 願 平1-42940

⑧ 出 願 平1(1989)2月27日

⑨ 発 明 者 小 柳 滋 神奈川県川崎市幸区小向東芝町1 株式会社東芝総合研究所内

⑩ 出 願 人 株 式 会 社 東 芝 神奈川県川崎市幸区堀川町72番地

⑪ 代 理 人 弁 理 士 則 近 憲 佑 外1名

明 細 書

1. 発明の名称

情報検索システム

2. 特許請求の範囲

(1) キーワードの付与されたデータの検索を行なう情報検索システムにおいて、キーワード間の関連を表わすキーワード相関テーブルを備え、データを格納するときは付与されたキーワードをキーワード相関テーブルを用いて変換して得られた特徴ベクトルをデータに付与し、検索するときは検索用のキーワードをキーワード相関テーブルを用いて変換して得られた特徴ベクトルを用いて検索を行なうことを特徴とする情報検索システム。

(2) キーワードの付与されたデータの検索を行なう情報検索システムにおいて、キーワード間の関連を表わすキーワード相関テーブルと、二層から成るニューラルネットワークを備え、データを格納するときは付与されたキーワードをキーワード相関テーブルを用いて変換して得られた特徴ベクトルをニューラルネットの結合係数として格納し、

検索するときは検索用のキーワードをキーワード相関テーブルを用いて変換して得られた特徴ベクトルをニューラルネットに入力することにより検索を行なうことを特徴とする情報検索システム。

3. 発明の詳細な説明

〔発明の目的〕

(産業上の利用分野)

本発明は、ニューラルネットワークを用いて情報検索を行なうための情報検索システムに関する。

(従来の技術)

最近、人間の右脳に似た情報処理能力を実現する方法としてニューラルネットワークが注目されている。ニューラルネットワークでは神経細胞をモデルとし、入力信号の重みつき総和をとり、これにしきい値作用と非線型関数を作用させることにより出力を生成するニューロンが多数個結合されたものである。

ニューラルネットワークを用いて情報検索システムを構築できれば、人間の右脳に似た連想、推測等の機能が利用でき、従来の情報検索システム

を高度化することが期待される。しかるに、現状では、単純な連想メモリが実現されているだけであり、情報検索システムとして利用できるレベルには到達していない。

従来の情報検索システムでは、検索手段としてキーワードの論理式による表現を用いている。またキーワードを体系化したシソーラスを備え、与えられたキーワードの下位概念を自動的に加えて検索することも可能である。しかるに、ユーザの検索したい内容を人間の右脳機能に相当するキーワードの論理式のみで表現するのは困難であり、またこれを支援するための適切なシソーラスを構築するのは困難である。

#### (発明が解決しようとする課題)

以上のように従来の情報検索システムではキーワードの論理式やシソーラスを用いた検索手段が提供されているが、より柔軟な検索を可能とするために、右脳的な検索手段をとり入れた情報検索システムの実現が望まれている。

本発明は、上記の課題を実現するためになされ

共に、ニューラルネットワーク内に格納される。すなわち、個々の格納データの特徴ベクトルは、格納データに対応するニューロンの重みベクトルとして保持される。

また、検索時には、検索したいキーワードをキーワード相関テーブルを用いて変換された検索ベクトルを用い、これと各格納データの特徴ベクトルとの類似度をニューラルネットワークにより求める。

これにより、検索は単なるキーワードマッチング以上に、キーワード間の相関関係を用いた機能を利用することができ、柔軟性に富んだ検索機能が実現できる。

#### (実施例)

以下、本発明の詳細を一実施例により説明する。

第1図は本発明の一実施例のシステム構成を説明する。

ここでは文献検索システムを例にとり説明する。各文献毎にあらかじめキーワードが付与されるとする。このキーワードの集合について各々の

たものであり、ニューラルネットワークを用いて情報検索システムにおける柔軟な検索手段を提案することを目的としている。

#### (発明の構成)

##### (課題を解決するための手段)

本発明では、情報検索システムで用いる全てのキーワードについて、その相関関係をあらかじめ計算してキーワード相関テーブルを作成する。

また、ニューラルネットワークは、個々のキーワードを表わすニューロンから構成される層と、個々の格納データを表わすニューロンから構成される層の2層から成る。この2層間は完全結合される。

各ニューロンは入力ベクトルと重みベクトルとの積和演算(内積)を行ない、その結果がしきい値を越えたものが発火する。

#### (作用)

本発明によれば、データベース内に格納されるデータは、それに対応するキーワードをキーワード相関テーブルにより変換された特徴ベクトルと

相関関係をキーワード相関テーブル10に格納する。キーワードを全部でA-Gの7種類としたときのキーワード相関テーブル10の一例を第2図に示す。キーワード相関テーブルの作り方は各種の方法が考えられるが、ここでは、文献内のキーワードの出現頻度に基づき、次のように定義する。テーブルの*i*行*j*列を $K_{ij}$ とすると

$$K_{ij} = F(i, j) / F(i)$$

但し、 $F(i)$ はキーワード*i*の出現回数、 $F(i, j)$ はキーワード*i*と*j*が同一文献内に同時に出現する回数を表わす。

このキーワード相関テーブルを用いて、文献データをニューラルネットワークに格納する方法について述べる。格納用特徴ベクトル生成部11は、各文献に付与されたキーワードより、キーワード相関テーブル10を用いてニューラルネットワーク12の重みを設定する。ニューラルネットワークの構成図を第3図に示す。第3図に示すようにニューラルネットワークは各キーワードを表わす層と、各文献を表わす層間の完全結合網とする。

次に重みの設定法について説明する。ここで文献P1はキーワードA, Bをもつとする。このとき、第2図のキーワード相関テーブルよりA行とB行の内容を参照する。

$$A = (1 \ 0.1 \ 0 \ 0 \ 0.1 \ 0.9 \ 0.8)$$

$$B = (0.2 \ 1 \ 0 \ 0.4 \ 0.3 \ 0.3 \ 0)$$

文献P1の特徴ベクトルP1はAとBの平均をとることにより生成する。但し、いずれかの要素が1のときは平均をとらずに結果は1とする。これによりP1は

$$P1 = (1 \ 1 \ 0 \ 0.2 \ 0.2 \ 0.6 \ 0.4)$$

となる。同様に文献P2がキーワードC, Eをもつとすると

$$P2 = (0.1 \ 0.1 \ 1 \ 0.4 \ 1 \ 0 \ 0.1)$$

文献P3がキーワードF, Gをもつとすると

$$P3 = (0.6 \ 0.1 \ 0 \ 0.1 \ 0.2 \ 1 \ 1)$$

となる。このようにして、格納する文献の特徴ベクトルP1, P2, P3が生成できる。ここで、文献P1に対応するニューロンの重みは特徴ベクトルP1に基づいて第4図に示すように設定する。

$$f(x) = 1 : x \geq 0 \text{ のとき}$$

$$= 0 : x < 0 \text{ のとき}$$

とすると、P1, P2, P3に対応するニューロンは次の演算を行なう

$$P1 : f(P1 \cdot q - 1)$$

$$= f((1 \ 1 \ 0 \ 0.2 \ 0.2 \ 0.6 \ 0.4)$$

$$\cdot (1 \ 0.1 \ 0 \ 0 \ 0.1 \ 0.9 \ 0.8) - 1)$$

$$= f(1 + 0.1 + 0.02 + 0.54 + 0.32 - 1) = f(0.98) = 1$$

$$P2 : f(P2 \cdot q - 1)$$

$$= f((0.1 \ 0.1 \ 1 \ 0.4 \ 1 \ 0 \ 0.1)$$

$$\cdot (1 \ 0.1 \ 0 \ 0 \ 0.1 \ 0.9 \ 0.8) - 1)$$

$$= f(-0.71) = 0$$

$$P3 : f(P3 \cdot q - 1)$$

$$= f((0.6 \ 0.1 \ 0 \ 0.1 \ 0.2 \ 1 \ 1)$$

$$\cdot (1 \ 0.1 \ 0 \ 0 \ 0.1 \ 0.9 \ 0.8) - 1)$$

$$= f(1.33) = 1$$

以上により、キーワードAの検索要求に対し、文献P1とP3に対応するニューロンが発火する。ここで文献P1はキーワードAとBをもち、文献P3はキーワードFとGをもつ。明らかに正しい

すなわち、各文献に対応するニューロンは特徴ベクトルの値を重みとするものとする。各ニューロンP<sub>i</sub>の演算は入力ベクトルをxとすると

$$f(P_i \cdot x - \theta)$$

で表わされる。但し、 $\theta$ はしきい値、 $f$ は非線型関数とする。

このようにしてニューラルネットワークを構成することができる。次に検索について述べる。

今、キーワードAをもつ文献について検索したいものとする。第1図の検索用特徴ベクトル生成部13により、検索要求に対応する特徴ベクトルが生成される。ここではキーワードAに対応する行をキーワード相関テーブル10より参照することにより、検索用特徴ベクトルqは、第2図のキーワード相関テーブルを用いると

$$q = (1 \ 0.1 \ 0 \ 0 \ 0.1 \ 0.9 \ 0.8)$$

となる。この検索用特徴ベクトルをニューラルネットワーク12の入力とすることにより検索が行なわれる。ここで各ニューロンの演算のしきい値 $\theta = 1$ 、非線型関数 $f$ を

キーワードを含むものは正しく検索され、またキーワードAをもたない文献でもそれと関連の深いキーワードをもつならば検索することができる。

このように、ニューラルネットワークを利用することにより、個々のキーワードの照合だけでなく、与えられたキーワード全体の中から判断することができるため、柔軟な検索が可能となる。

なお、本発明はニューラルネットワークを利用しなくても通常の計算機上で同様の処理を行なうことにより同じ効果が得られることは明らかである。

#### 〔発明の効果〕

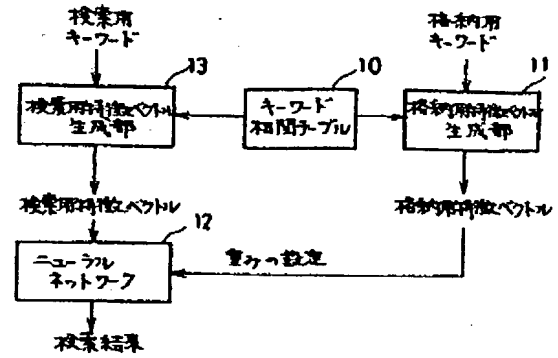
以上に述べたように、本発明によれば情報検索システムにおいてニューラルネットワークを利用することにより、単なるキーワードの照合だけでなく、キーワード間の相関関係に基づいた検索が可能となるため、従来よりも高機能の検索を実現することができる。

#### 4. 図面の簡単な説明

第1図は本発明の一実施例のシステム構成を表

わす図、第2図はキーワード相関テーブルの内容を表わす図、第3図はニューラルネットワークの構成図、第4図は文献P1に対応するニューロンの重みを表わす図である。

- 10…キーワード相関テーブル
- 11…格納用特徴ベクトル生成部
- 12…ニューラルネットワーク
- 13…検索用特徴ベクトル生成部



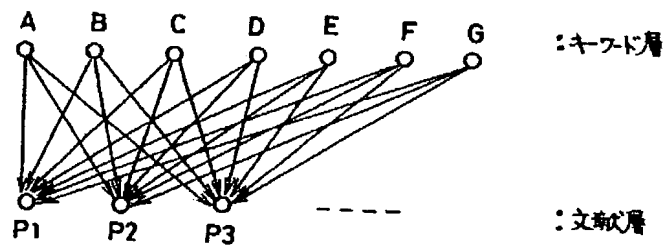
第 1 図

代理人 井理士 則 近 憲 佑

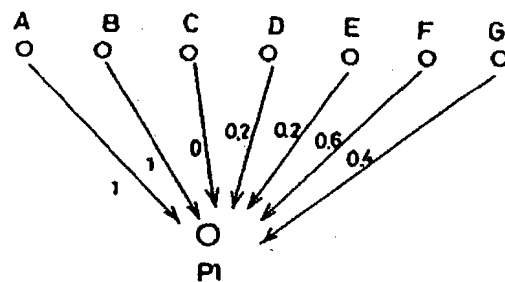
岡 松 山 允 之

	A	B	C	D	E	F	G
A	1	0.1	0	0	0.1	0.9	0.8
B	0.2	1	0	0.4	0.3	0.3	0
C	0	0.1	1	0	0.6	0	0
D	0	0.3	0	1	0.9	0	0.1
E	0.2	0.1	0.8	0.8	1	0	0.2
F	0.6	0.2	0	0	0	1	0.8
G	0.6	0	0	0.2	0.4	0.7	1

第 2 図



第 3 図



第 4 図